ORIGINAL ARTICLE

Metabolite Profiles of the Serum of Patients with Non–Small Cell Carcinoma

Peter J. Mazzone, MD, FCCP, MPH,[a,*] Xiao-Feng Wang, PhD,[b] Mary Beukemann,[a] Qi Zhang, MS,[b] Meredith Seeley,[a] Rob Mohney, PhD,[c] Tracy Holt, PhD,[c] Kirk L. Pappan, PhD[c]

[a]*Pulmonary Department, Cleveland Clinic, Cleveland, OH, USA*
[b]*Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA*
[c]*Metabolon, Durham, NC, USA*

Received 28 April 2015; revised 10 September 2015; accepted 11 September 2015

ABSTRACT

Introduction: Alterations of serum metabolites may allow us to identify individuals with lung cancer and advance our understanding of the nature and treatment of their cancer. We aimed to identify serum metabolites that differentiate patients with lung cancer from at-risk controls.

Methods: Serum samples from patients with biopsy-confirmed untreated stage I through stage III non–small cell lung cancer and at-risk controls were divided into fractions for analysis by ultrahigh-performance liquid chromatography–tandem mass spectrometry and gas chromatography–mass spectrometry. Compounds were identified by comparison with library entries of purified standards. Differences in concentrations of single metabolites and metabolite ratios were identified. Prediction models were developed.

Results: Serum samples from 284 subjects was analyzed. The subjects' mean age was 67 and 48% were female. Ninety-four patients had lung cancer (50 had adenocarcinoma and 44 had squamous cell carcinoma), 44% had stage I disease, 17% had stage II disease, and 39% had stage III disease. The patients with cancer were slightly older than the controls (68.7 versus 66.2 years, $p = 0.013$). A total of 534 metabolites were identified in eight metabolite super-pathways and 73 subpathways. The concentrations of 149 metabolites differed significantly (q values <0.05) between the cancer and control groups (70 were lower in the cancer group and 79 were higher), and 9723 metabolite ratios differed significantly (q values <0.001) between the cancer and control groups. The accuracies of the models (cancer and cancer subtypes versus control) trained on 70% of the subjects and tested on 30% (expressed as C-statistics) ranged from 0.748 to 0.858.

Conclusions: Differences in the serum metabolite profile exist between patients with stage I through stage III non–small cell lung cancer and matched controls.

Keywords: Adenocarcinoma; Squamous cell carcinoma; Metabolites; Biomarkers

Introduction

The pathogenesis of lung cancer involves a complicated interplay between our natural defenses and the effect of exposure to extrinsic insults capable of altering the normal cellular life cycle. Uncorrected modifications in cellular behavior can be identified as molecular changes across the spectrum of systems biology. Alterations in genetic, transcriptional, and proteomic function result in changes in the metabolic properties of cancer cells. These metabolic changes are reflected in the composition of small molecule metabolites in cancer tissue, blood, and urine. Knowledge of distinguishing patterns of small molecule metabolites may allow us to identify individuals with lung cancer, better understand the nature of their cancer, and develop novel means of targeting the cancer.

Metabolite profiling is a relatively underrepresented field of biomarker development for identifying and

*Corresponding author.

Disclosure: The authors declare no conflict of interest.

An abstract of this article was published previously in the journal *Chest* (2014;146:587A).

Address for correspondence: Peter J. Mazzone, MD, FCCP, MPH, Respiratory Institute, Cleveland Clinic, 9500 Euclid Ave., A90, Cleveland, OH 44195. E-mail: mazzonp@ccf.org

ISSN: 1556-0864

http://dx.doi.org/10.1016/j.jtho.2015.09.002

characterizing lung cancer. Metabolic changes can be identified as differences in the concentration of a single metabolite or alterations of the constituents of an entire metabolic pathway. Previous studies have identified metabolite differences in the blood of patients with non–small cell cancer that are related to altered amino acid metabolism, glycolysis and gluconeogenesis, protein metabolism, handling of oxidative stresses, and fatty acid metabolism.[1–12] Metabolite profiles have made it possible to distinguish patients with lung cancer from those without.[1–12]

Lung cancer is a heterogeneous disease that develops in a heterogeneous population.[13] The metabolic profile of each individual is somewhat unique, as is the metabolic behavior of his or her cancer.[14] This uniqueness makes it difficult to identify one distinct metabolic profile of lung cancer. Global metabolite profiling may provide a broad overview of the potential metabolic alterations, thereby allowing for refinements in profiling to occur on the basis of the phenotype of the patient and the cancer. In this study, we used global metabolite profiling of serum with the aim of identifying metabolic characteristics that differentiated patients with stage I through stage III primary lung adenocarcinoma and squamous cell carcinoma from a phenotypically similar group of patients without lung cancer. Some of the results of this study have been reported previously in the form of an abstract.[15]

Materials and methods

The study protocol was approved by the Institutional Review Board of the Cleveland Clinic (10-521). All study subjects signed written informed consent. Metabolomics profiling services were performed by Metablon under contract with the Cleveland Clinic. The Cleveland Clinic paid for these services and owns the data presented. The results of metabolite profiling were received by the Cleveland Clinic before Metabolon received information about the category of the samples (cancer or control). All statistical analyses presented in the manuscript were performed by X. F. W. and Q. Z. of the Cleveland Clinic.

Study subjects

All samples from the subjects were obtained from a lung cancer biorepository maintained by the Cleveland Clinic. Study subjects were included in the biorepository if they had biopsy-confirmed untreated lung cancer or if they were at risk for development of lung cancer on the basis of an age older than 40 years and tobacco use of at least 10 pack-years, a family history of lung cancer, or the presence of chronic obstructive pulmonary disease. Study subjects were excluded from the biorepository if they had a previous history of lung cancer, had a history of another cancer within 5 years, were receiving immunosuppression, or were using continuous supplemental oxygen. For this study, all patients in the cancer group had adenocarcinoma or squamous cell carcinoma and stage I through stage III disease. Two control subjects were selected for each patient with cancer through propensity matching, with age, sex, smoking history, presence or absence of diabetes, elevated cholesterol level, and presence of chronic obstructive pulmonary disease used for the matching. Data collection included demographic variables and comorbidities for all subjects, histologic diagnosis of cancer, stage, and survival for the subjects in the cancer group. All blood samples were drawn and stored at −80°C as 500-μL aliquots of serum until they were sent to Metabolon on dry ice as batched, de-identified samples for testing. All samples were maintained at Metabolon at −80°C until processed.

Sample preparation

To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and recover chemically diverse metabolites, the proteins were precipitated with methanol by vigorous shaking for 2 minutes (Glen Mills Inc., Clifton, NJ) as described previously.[16,17] The resultant extract was divided into four fractions: one for analysis by ultrahigh-performance liquid chromatography–tandem mass spectrometry (UPLC-MS/MS) run in positive mode, one for analysis by UPLC-MS/MS run in negative mode, one for analysis by gas chromatography–mass spectrometry (GC/MS), and one aliquot that was retained for backup analysis, if needed.

Mass spectrometry analysis

Nontargeted UPLC-MS/MS and GC/MS analyses were performed at Metabolon, Inc., as described.[16–18] The UPLC-MS/MS portion of the platform incorporates a Waters Acquity UPLC system (Waters Corporation) and a Thermo-Finnigan LTQ mass spectrometer (Thermo Fisher Scientific Inc.), including an electrospray ionization source and linear ion-trap mass analyzer. Aliquots of the vacuum-dried sample were reconstituted, one each in acidic or basic liquid chromatography–compatible solvents containing eight or more injection standards at fixed concentrations (to ensure both injection and chromatographic consistency). Extracts were loaded onto columns (Waters UPLC BEH C18-2.1 x 100 mm, 1.7 μm, Waters Corporation) and gradient eluted with water and 95% methanol containing 0.1% formic acid (acidic extracts) or 6.5 mM ammonium bicarbonate (basic extracts). The instrument was set to

scan 99 to 1000 m/z and alternated between MS and MS/MS scans.

Samples analyzed by GC/MS were dried under vacuum desiccation for a minimum of 18 hours before being derivatized by using bis(trimethylsilyl)trifluoroacetamide as described.[19] Derivatized samples were separated on a 5% phenyldimethyl silicone column with helium as the carrier gas and a temperature ramp from 60° to 340°C within a 17-min period. All samples were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer (Thermo Fisher Scientific Inc.) operated at unit mass resolving power with electron impact ionization and a scan range of 50 to 750 atomic mass units.

Quality control

All the columns and reagents to complete all related experiments were purchased in bulk from a single lot. For monitoring of data quality and process variation, multiple replicates of extracts from a pool of human plasma were prepared in parallel and injected throughout the run, interspersed among the experimental samples. Instrument variability was determined by calculating the median relative standard deviation for the standards that were added to each sample before injection into the mass spectrometers (median relative standard deviation 6%). Overall process variability was determined by calculating the median relative standard deviation for all endogenous metabolites (i.e., noninstrument standards) present in 100% of the pooled human plasma samples (median relative standard deviation 12%). In addition, process blanks and other quality control samples were spaced evenly among the injections for each day, and all experimental samples were randomly distributed throughout each day's run.

Compound identification, quantification, and data curation

Metabolites were identified by automated comparison of the ion features in the experimental samples to a reference library of chemical standard entries that included retention time, molecular weight, preferred adducts, and in-source fragments as well as associated MS spectra and curated by visual inspection for quality control using software developed at Metabolon.[20] Identification of known chemical entities was based on comparison with metabolomic library entries of more than 3500 commercially available purified standards. Peaks were quantified using area under the curve. Raw area counts for each metabolite in each sample were normalized to correct for variation resulting from interday differences in instrument tuning by the median value for each run-day, with the medians set to 1.0 for each run. Missing values were imputed with the observed minimum after normalization.

Statistical methods

Cancer cases were matched to controls (see earlier) on the basis of the results of propensity score matching using the greedy matching technique to reduce sample selection bias. Demographic variables were described by using sample mean with standard deviation or proportion as appropriate. The study group was divided into cancer and control groups. Categorical variables were compared using the Pearson chi-square test, whereas continuous variables were compared using the two-sample independent t test.

Metabolite concentrations were returned to the Cleveland Clinic as raw data and scaled data for analysis. Missing values in the data set, which were generally due to the metabolite level falling below the instrument's limit of detection, were imputed with the minimum measured value of that metabolite to account for the missing data. Statistical analyses were performed on natural log-transformed data. Differences in metabolite concentrations between the cancer and control groups, as well as (separately) differences in ratios of metabolite concentrations, were first assessed with a univariate robust t test for each metabolite using the method of empirical Bayesian shrinkage of the standard errors,[21] followed by the application of Benjamini and Hochberg's multiple comparison approach to determination of the false discovery rate.[22] This comparison was performed first for the entire group of subjects with cancer and control subjects and then separately by histologic diagnosis.

We built all the models presented by using the full data sets and including all metabolites. Our procedure for building the prediction model included two additional steps. First, a resample-based feature elimination algorithm that incorporated correlation analysis was conducted to select the important features and avoid multicollinearity of model fitting.[23] Second, random forest models were built using the selected variables to predict cancer. Models were developed for all subjects with lung cancer and all controls and then separately on the basis of histologic diagnosis. Models were built using the entire data set for each comparison and separately on the basis of a training set of 70% of the subjects with testing on the remaining 30%. This process was repeated 100 times to avoid randomness of data splitting. Drugs and drug metabolites were excluded from the model-building procedure (72 in total, 10 of which differed between the cancer and control groups) and from the metabolite ratio calculations. C-statistics were calculated by the rank correlation between predicted

probabilities of the outcome and the observed response.

We first determined the accuracy of the model of cancer versus control from differences in concentrations of single metabolites and then reexamined the accuracy using metabolite ratios to see whether a potential substrate–product pairing or otherwise unknown metabolite relationship might strengthen the model in comparison to individual metabolites. Ratios of normalized concentrations of each single metabolite to all other single metabolites, excluding xenobiotics, were used in this analysis. All analyses were performed by using the R statistical package (www.r-project.org). Additional detail on the methods used in this study is provided in the online data supplement.

Results

Serum samples from 284 subjects were analyzed. Their mean age was 67 years, and 48% of them were female. Ninety-four patients had lung cancer (50 had adenocarcinoma and 44 had squamous cell carcinoma); 44% had stage I disease, 17% had stage II disease, and 39% had stage III disease. The patients with cancer were slightly older than the controls (68.7 versus 66.2 years, $p = 0.013$); all other matched variables were not significantly different (Table 1).

A total of 534 metabolites were identified in eight metabolite superpathways and 73 subpathways. After correction for multiple comparisons ($q < 0.05$), the concentration of 149 metabolites differed significantly between the cancer and control groups (70 lower in the cancer group and 79 higher) (Table 2 and Supplementary Tables 1 and 2). Sixty-five metabolite concentrations differed significantly between the subjects with adenocarcinoma and matched controls after adjustment for multiple comparisons ($q < 0.05$), five of which were unique (i.e., not different in the aforementioned comparison of all subjects with cancer and control subjects).

Table 1. Demographics, comorbidities, and cancer characteristics of the study cohort

Characteristic	Patients with cancer (n = 94)	Controls (n = 190)	p Value
Age (y)	68.7	66.2	0.0125
Female sex (%)	44.7	49.5	0.4468
History of smoking (%)	95.7	97.4	0.4623
COPD (%)	29.8	23.7	0.2681
DM (%)	11.7	8.9	0.4637
Elevated lipid level (%)	21.3	19.5	0.7211
Adenocarcinoma (n)	50		
Squamous cell (n)	44		
Stage I (%)	44		

COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus.

Table 2. Metabolite differences categorized by metabolite superpathway

Superpathway	Number	Different	Up[a]	Down[a]
Amino acid	96	22	10	12
Peptide	30	12	5	7
Carbohydrate	25	4	3	1
Energy	9	0	0	0
Lipid	191	68	44	24
Nucleotide	22	5	4	1
Cofactors and vitamins	25	8	3	5
Xenobiotics	136	30	10	20
Total	534	149	79	70

[a]Up indicates a higher level in the patients with cancer than in the controls, and down indicates a lower level in the patients with cancer than in the controls.

Fifty metabolite concentrations differed significantly between subjects with squamous cell carcinoma and matched controls after adjustment for multiple comparisons ($q < 0.05$), including three that were unique. None of the metabolite concentrations differed between subjects with adenocarcinoma and squamous cell carcinoma after adjustment for multiple comparisons. Of the 534 metabolites, 72 were classified as drugs or drug metabolites. Most of these metabolites were present in fewer than 10% of the study subjects. Those present in greater than 10% of study subjects included acetaminophen, nonsteroidal antiinflammatory drugs, and their metabolites. The concentration of 10 of the 72 differed between the cancer and control groups; nine were elevated (eight were acetaminophen and its metabolites), and one was lower. Three lipid-lowering drugs were identified—none in more 10% of study subjects and none with different levels between the cancer group and the controls.

Metabolite ratios were evaluated separately. Of the 106,491 metabolite ratios (after correction for multiple comparisons at q values less than 0.05, less than 0.01, and less than 0.001, respectively), 34,680, 20,196, and 9723 differed between all subjects with cancer and control subjects (the top 64 were used for model building, see Supplementary Table 3); 17,985, 6648, and 1249 differed significantly between subjects with adenocarcinoma and matched controls the (top 57 were used for model building); and 16,917, 5921, and 1390 differed significantly between subjects with squamous cell carcinoma and matched controls (the top 38 were used for model building). No metabolite ratios differed significantly between subjects with adenocarcinoma and squamous cell carcinoma after adjustment for multiple comparisons.

Models using metabolite concentrations and separate models using metabolite concentration ratios made it possible to separate cancer groups from control groups. The accuracies (expressed as C-statistics) of models built

Table 3. Accuracy of models of cancer versus control built from differences in concentrations of single metabolites and metabolite concentration ratios

	Single-metabolite models			Metabolite ratio models		
Type of cancer	Model C-statistic	Model OOB	Validation C-statistic (CI)	Model C-statistic	Model OOB	Validation C-statistic (CI)
Lung cancer	0.869	0.193	0.831 (0.823-0.840)	0.908	0.183	0.848 (0.839-0.858)
Adenocarcinoma	0.814	0.211	0.748 (0.734-0.761)	0.938	0.165	0.753 (0.739-0.768)
Squamous cell carcinoma	0.926	0.144	0.839 (0.827-0.850)	0.955	0.121	0.858 (0.844-0.873)

OOB, out of bootstrap; CI, confidence interval.

on the entire data set ranged from 0.814 to 0.967, whereas the validated accuracies of models trained on 70% of the subjects and tested on 30% ranged from 0.748 to 0.858 (Table 3). The single-metabolite models included 20 metabolites each (Table 4).

Discussion

We report the results of global small molecule metabolite profiling of the serum of patients with stage I through stage III adenocarcinoma and squamous cell carcinoma. Metabolite concentrations and ratios of metabolite concentrations in the patients with cancer were compared with those of controls who were matched for demographic and comorbidity characteristics that could alter their metabolite profile. We identified 149 metabolites that were present in different concentrations in the subjects with lung cancer. Metabolite differences were identified across multiple metabolite superpathways. Models developed on the basis of these differences made it possible to separate subjects with lung cancer from control subjects with good accuracy.

Several previous studies have evaluated blood metabolite profiles of patients with lung cancer.[1–12] The studies have differed in terms of the breadth of metabolite profiling (a single superpathway versus global profiling), the profiling methodology used, and the populations studied. These studies have suggested differences in amino acid metabolism, glycolysis and gluconeogenesis, protein metabolism, handling of oxidative stresses, and fatty acid metabolism in the population with lung cancer. Our study supports the presence of these global alterations in metabolism. In addition, our results suggest that in those with lung cancer, phenolic compounds are reduced and activity of the transsulfuration pathway is increased.

Although our results globally support what has been reported in the literature, the specific metabolite alterations identified in our study are different from those identified by other investigators. Some of the differences in metabolite concentration we report (e.g., for histidine and tryptophan) are concordant with previous reports, a few (e.g., for sarcosine) are discordant, others have variable results in the literature (high in one report and low in another), and several metabolites that have previously been reported to be present in different concentrations in patients with lung cancer were not different in our study.[1–12] Most of the specific metabolite differences that we report have not appeared in other studies. This could be due to differences in study methodology and the populations studied, or it could simply reflect a lack of depth of analysis of the route metabolite pathway alterations in the current study. Further analysis of route metabolite pathway alterations may tie together the findings of the current and previous studies. No previous lung cancer study has evaluated differences in the ratios of metabolite concentrations.

A deeper understanding of the metabolic pathways that are altered to produce the concentration differences that we have identified would help in validating and applying our findings. An atlas of genetic influences on human blood metabolites has been developed.[24] Thirty of the 149 metabolites whose concentrations differed between patients with lung cancer and the controls in our study have been connected with the potential influence of a single nucleotide polymorphism.[24] In our review of the literature, none of these 30 single-nucleotide polymorphisms have been reported to be associated with lung cancer.

Our study has several strengths. We narrowed our population of patients with cancer to those with stage I through stage III adenocarcinoma and squamous cell carcinoma, thus minimizing the influence of lung cancer heterogeneity and advanced disease on the results. We used propensity matching to identify controls whose demographics and comorbidities were similar to those of the subjects with lung cancer, which minimized these influences on the metabolite profiles. We used technically validated methodologies that incorporate extensive quality controls. We used conservative statistical analysis techniques, were careful to control for the multiple comparisons in this large data set, and analyzed

Table 4. Metabolites within the models of single metabolites

Cancer vs. control	Adenocarcinoma vs. control	Squamous cell vs. control
Arachidonate	Arachidonate	Pelargonate
Phosphate	Phosphate	Methionine
Pelargonate	Pelargonate	*N*1-Methyladenosine
Tyrosine	Taurodeoxycholate	*N*-Acetylmethionine
Methionine	Nonadecanoate	*N*-Acetylneuraminate
Nonadecanoate	Cortisol	Ascorbate
Inositol 1-phosphate	Quinate	Azelate
α-Tocopherol	Theobromine	Hypoxanthine
Hippurate	Glycocholate	Cysteine
N-Acetylalanine	Stearoyl sphingomyelin	*N*-(2-Furoyl)glycine
N-Acetylmethionine	2-Hydroxybutyrate	DSGEGDFXAEGGGVR
N-Acetylneuraminate	3-Methyl-2-oxobutyrate	Acetylcarnitine
Ascorbate	1,6-Anhydroglucose	Hexanoylcarnitine
Cortisol	3-Hydroxydecanoate	Alanine
2-Hydroxystearate	Palmitoylcarnitine	δ-Tocopherol
Paraxanthine	Hypoxanthine	γ-Tocopherol
Riboflavin	Cysteine	Eicosenoate
Quinate	*N*-(2-Furoyl)glycine	ADpSGEGDFXAEGGGVR
Azelate	Acetylcarnitine	1-Methylxanthine
Theobromine	Glycochenodeoxycholate	Laurylcarnitine

metabolite ratios separately, thereby highlighting the interdependence of metabolites. Weaknesses of our study include an incomplete ability to connect the metabolite differences to specific defined metabolic pathways and validate our findings in separate patient cohorts. This fact is particularly important because lack of external validation of metabolite findings has been a concern of metabolomics research. Not all potential influences on the metabolite profiles (e.g., obesity) could be controlled for. Addressing these weaknesses is a goal for the future.

In conclusion, differences in serum metabolite profiles exist between patients with stage I through stage III non–small cell carcinoma and matched controls. A deeper understanding of these differences could provide us with insight into the pathobiology of lung cancer, lead to the development of novel lung cancer biomarkers, and identify new targets for therapy.

Acknowledgments

This project was made possible by the generous support of a gift from the William C. Meier Foundation. We would like to acknowledge the Clinical Research Unit of the Cleveland Clinic for its support in processing, storing, and shipping the serum specimens. P. J. M. takes responsibility for the content of the manuscript, including the data and analysis. P. J. M. made substantial contributions to the conception and design of the research, acquisition of data, analysis and interpretation of data, drafting and revision of the article, and he provided final approval of the version to be published. X. W. and Q. Z. made substantial contributions to the analysis and interpretation of data and revision of the submitted article, and they have provided final approval of the version to be published and agreed to be accountable for all aspects of the work. R.M., T.H., and K.P. made substantial contributions to the conception and design of the research, acquisition of data, analysis and interpretation of data, and revision of the manuscript, and they have provided approval of the version submitted and agree to be accountable for all aspects of the work. M.B. and M.S. made substantial contributions to the acquisition of data and revision of the submitted article, provided final approval of the version submitted, and agree to be accountable for all aspects of the work. All the statistical analysis reported was performed by the Cleveland Clinic.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at http://dx.doi.org/10.1016/j.jtho.2015.09.002.

References

1. Maeda J, Higashiyama M, Imaizumi A, et al. Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study. *BMC Cancer*. 2010; 10:690.
2. Miyagi Y, Higashiyama M, Gochi A, et al. Plasma free amino acid profiling of five types of cancer patients and

its application for early detection. *PLoS ONE*. 2011; 6:e24143.
3. Hori S, Nishiumi S, Kobayashi K, et al. A metabolomic approach to lung cancer. *Lung Cancer*. 2011;74:284–292.
4. Rocha CM, Carrola J, Barros AS, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. *J Proteome Res*. 2011; 10:4314–4324.
5. Wu M, Xu Y, Fitch WL, et al. Liquid chromatography/mass spectrometry methods for measuring dipeptide abundance in non-small-cell lung cancer. *Rapid Commun Mass Spectrom*. 2013;27:2091–2098.
6. Wen T, Gao L, Wen Z, et al. Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. *Mol Biosyst*. 2013;9:2370–2378.
7. Fan TWM, Lane AN, Higashi RM, et al. Altered regulation of metabolic pathways in human lung cancer discerned by ^{13}C stable isotope-resolved metabolomics (SIRM). *Mol Cancer*. 2009;8:41.
8. Shingyoji M, Iizasa T, Higashiyama M, et al. The significance and robustness of a plasma free amino acid (PFAA) profile-based multiplex function for detecting lung cancer. *BMC Cancer*. 2013;13:77.
9. Dong J, Cai X, Zhao L, et al. Lysophosphatidylcholine profiling of plasma: discrimination of isomers and discovery of lung cancer biomarkers. *Metabolomics*. 2010;6:478–488.
10. Guo Y, Wang X, Qiu L, et al. Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using Fourier transform ion cyclotron resonance mass spectrometry. *Clinica Chimica Acta*. 2012;414:135–141.
11. Lokhov PG, Kharybin ON, Archakov AI. Diagnosis of lung cancer based on direct-infusion electrospray mass spectrometry of blood plasma metabolites. *Int J Mass Spectrom*. 2012;309:200–205.
12. Liu J, Mazzone PJ, Cata JP, et al. Serum free fatty acid biomarkers of lung cancer. *Chest*. 2014;146:670–679.
13. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small cell lung cancers: a heterogeneous set of diseases. *Nature Rev*. 2014;14:535–546.
14. Kusano M, Mendez E, Furton KG. Comparison of the volatile organic compounds from different biological specimens for profiling potential. *J Forensic Sci*. 2013; 58:29–39.
15. Mazzone PJ, Wang X, Beukemann M, et al. The analysis of small molecule metabolite profiles in the blood as a biomarker of lung cancer. *Chest*. 2014;146:587A.
16. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem*. 2009;81:6656–6667.
17. Weiner J 3rd, Parida SK, Maertzdorf J, et al. Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS ONE*. 2012;7:e40221.
18. Sha W, da Costa KA, Fischer LM, et al. Metabolomic profiling can predict which humans will develop liver dysfunction when deprived of dietary choline. *FASEB J*. 2010;24:2962–2975.
19. Ohta T, Masutomi N, Tsutsui N, et al. Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. *Toxicol Pathol*. 2009;37:521–535.
20. DeHaven C, Evans A, Dai H, Lawton K. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J Cheminform*. 2010;2:9.
21. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Bio*. 2004;3: Article 3.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stast Methodol*. 1995;57:289–300.
23. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28:1–26.
24. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543–550.